ON THE MULTIVARIATE ANALYSIS

OF VARIANCE

By

DEL LON WEST

Bachelor of Science

Southeastern State College

Durant, Oklahoma

1957

Submitted to the faculty of the Graduate School of
the Oklahoma State University in partial
fulfillment of the requirements
for the degree of
MASTER OF SCIENCE
1959

ON THE MULTIVARIATE ANALYSIS OF VARIANCE

Thesis Approved:

Franklin A. Graybill

Thesis Adviser

Carl E. Marshall

Robert MacVicar

Dean of the Graduate School

PREFACE

It is well known that the sums of squares in Eisenhart's model I are distributed as Chi-Squares.

In this thesis we shall prove some theorems in the multivariate case which are parallel to theorems already established in the univariate case. We shall show that in the multivariate analysis of variance table the forms are distributed as independent Wishart distributions.

The author is deeply indebted to Dr. Franklin A. Graybill for suggesting the problem and his guidance in the preparation of this thesis.

TABLE OF CONTENTS

Chapter		Page
I.	INTRODUCTION	1
	Necessary Theorems	2
II.	THEOREMS CONCERNING THE WISHART DISTRIBUTION	7
	Theorem 1	7
	Theorem 2	10
	Theorem 3	12
	Theorem 4	13
	Theorem 5	15
III.	APPLICATIONS	17
	The Sums of Squares in the Analysis of Variance Table	18
	Test Function	23
	Example	24
IV.	SUMMARY AND CONCLUSIONS	28
	Suggestions For Future Study	30
BIBLIOGRAPHY		31

LIST OF TABLES

Table		Page
I.	Analysis of Variance	19
II.	Data	25
III.	Covariance Analysis	26

CHAPTER I

INTRODUCTION

In this paper we shall use a capital letter with a subscript and a circumflex, $\underset{\sim}{Y}_i$, to denote a vector; a capital letter without a subscript, Y, to represent a matrix whose ij^{th} element is y_{ij}; and a capital letter with a subscript, an over dot, and a transpose mark, $\dot{\underset{\sim}{Y}}_i'$, to indicate a row from a matrix. j will denote a column vector where each element is equal to unity. J will denote a square matrix where each element is equal to unity. $\underset{\sim}{0}$ will denote a matrix, or vector, where each element is equal to zero, When we speak of an idempotent matrix, we shall imply a symmetric idempotent matrix, i.e., AA = A and A' = A. The symbol Σ will be used for the covariance matrix whose ij^{th} element is σ_{ij}, unless otherwise stated. The Wishart distribution will be denoted by $W(B \mid \Sigma, p, n)$ which indicates that the elements of the matrix B is distributed as Wishart with Σ as above, p components in the vectors, and sample size n. The frequency function of the Wishart distribution is

$$f(B) = \frac{|B|^{\frac{1}{2}(n-p-1)} \; e^{-\frac{1}{2}\operatorname{tr}(B\,\Sigma^{-1})}}{2^{\frac{1}{2}np}\; \pi^{p(p-1)/4}\; |\Sigma|^{n/2} \prod_{i=1}^{p} \Gamma[\frac{1}{2}(n+1-i)]}$$

where $tr(B\ \Sigma^{-1})$ means the trace of the product of the matrix B with the matrix Σ^{-1}.

Necessary Theorems

Theorem A. If A is an $(n \times n)$ symmetric matrix of rank p, then a necessary and sufficient condition that A is idempotent is that p of the characteristic roots of A are each equal to unity and the remaining $(n-p)$ characteristic roots are equal to zero. (3)

Theorem B. If A is an idempotent matrix, then the rank of A equals the trace of A. (3)

Theorem C. The only non singular idempotent matrix is the identity matrix. (3)

Theorem D. If A is an $(n \times n)$ idempotent matrix of rank p such that $p < n$ $(p = n)$, then A is a positive semidefinite matrix (positive definite matrix). (3)

Theorem E. If A and B are diagonal matrices such that $AB = 0$, then if the i^{th} diagonal element of A is non zero this implies that the i^{th} diagonal element of B is equal to zero. (3)

Theorem F. If A is an idempotent matrix whose i^{th} diagonal element is equal to zero, then every element in the i^{th} row and i^{th} column of A is equal to zero.

Theorem G. If B_i, $i = 1, 2, \ldots, q$, are a set of $(n \times n)$ symmetric matrices, then a necessary and sufficient condition that there exists an orthogonal matrix, R, such that RB_iR' are each diagonal is that $B_iB_j = B_jB_i$ for all i and j. (3)

Theorem H. Let B_i, i = 1, 2, . . . , q, be a collection of (n x n) symmetric matrices such that

$$\sum_{i=1}^{q} B_i = I_n.$$

Then any one of the conditions c_1, c_2, c_3, is necessary and sufficient for the remaining two.

c_1: Each B_i is an idempotent matrix.

c_2: $B_iB_j = \underset{\sim}{0}$ for all $i \neq j$.

c_3: $\sum_{i=1}^{q} n_i = n$ where n_i is the rank of B_i. (3)

Theorem I. If A is symmetric, then there exists an orthogonal matrix, R, such that $RAR' = D$ where D is a diagonal matrix whose i^{th} diagonal element is the i^{th} characteristic root of A. (3)

Theorem J. Suppose the p-component, independent vectors, $\underset{\sim}{Y}_i$, i = 1, 2, . . . , n, $(n \geq p)$ are each distributed as $N(\underset{\sim}{0}, \Sigma)$. Then

$$B = \sum_{i=1}^{n} \underset{\sim}{Y}_i \underset{\sim}{Y}_i'$$

is distributed as $W(B \mid \Sigma, p, n)$. (1)

Theorem K. Suppose the p-component vectors $\underset{\sim}{Y}_i$, i = 1, 2, . . . , n, $(n \geq p)$ are independent each with the distribution $N(\underset{\sim}{\mu}, \Sigma)$. Then the density of

$$B = \sum_{i=1}^{n} (\underset{\sim}{Y}_i - \overline{Y})(\underset{\sim}{Y}_i - \overline{Y})'$$

is $W(B \mid \not{\Sigma}, p, n-1)$ where

$$\overline{Y} = \sum_{i=1}^{n} \underset{\sim}{Y}_i. \quad (1)$$

Theorem L. Let

$$B = \sum_{i=1}^{n} \underset{\sim}{Y}_i \underset{\sim}{Y}_i'$$

where $\underset{\sim}{Y}_i$ is a vector of observations from $N(\underset{\sim}{\mu}_i, \not{\Sigma})$, and let

$$T = \sum_{i=1}^{n} \underset{\sim}{\mu}_i \underset{\sim}{\mu}_i'$$

also let k be the rank of T. Then the density of B is $W(B \mid \not{\Sigma}, p, n)$ times a function that depends on the roots of $|T - \lambda\not{\Sigma}| = 0$ and the roots of $|T - \lambda\not{\Sigma}B^{-1}\not{\Sigma}| = 0$. If $k = 1$, this function involves a Bessel function, $k = 2$, it involves an infinite series of Bessel functions, if $k = 3$, it can be expressed in terms of a triple integral. For higher values of k it is expressed as a multiple integral. Let us denote this function by $J_n(x)$. (1)

Theorem M. If the B_i, $i = 1, 2, \ldots, q$, are independently distributed according to $W(B_i \mid \not{\Sigma}, p, n_i)$ then

$$B = \sum_{i=1}^{q} B_i$$

is distributed according to $W(B \mid \Sigma, p, \sum_{i=1}^{q} n_i)$. (1)

The moment generating function of $W(B \mid \Sigma, p, n)$ is

$$m(t_{11}, t_{12}, \ldots, t_{pp}) = E\, e^{t_{11}a_{11} + t_{12}a_{12} + \cdots + t_{pp}a_{pp}}$$

$$= E\, e^{tr(tB)}$$

$$= \int_R \cdots \int_R \frac{|B|^{\frac{1}{2}(n-p-1)}\, e^{-\frac{1}{2} tr(\Sigma^{-1}B - 2BT)}}{K|\Sigma|^{n/2}}$$

$$= \frac{|\Sigma^{-1} - 2t|^{-\frac{n}{2}}}{|\Sigma|^{n/2}} \int_R \cdots \int_R \frac{|B|^{\frac{1}{2}(n-p-1)}\, e^{-\frac{1}{2} tr[(\Sigma^{-1} - 2t)B]}}{K\,|\Sigma^{-1} - 2t|^{-n/2}}$$

$$= \frac{1}{|\Sigma|^{n/2}\, |\Sigma^{-1} - 2t|^{n/2}}$$

$$= \frac{1}{|I - 2\Sigma t|^{n/2}}$$

where t is defined to be the (p x p) matrix,

$$t = \begin{bmatrix} t_{11} & \frac{1}{2}t_{12} & \cdots & \frac{1}{2}t_{1p} \\ \frac{1}{2}t_{12} & t_{22} & & \vdots \\ \vdots & & \ddots & \vdots \\ \frac{1}{2}t_{1p} & \cdots & \cdots & t_{pp} \end{bmatrix}$$

and where

$$K = 2^{n} \left|\Sigma\right|^{\frac{n}{2}} \pi^{\frac{1}{2}} \Gamma\left(\frac{n}{2}\right) \Gamma\left(\frac{n-1}{2}\right) .$$

CHAPTER II

THEOREMS CONCERNING THE WISHART DISTRIBUTION

<u>Theorem 1.</u> <u>If the p variate independent vectors, $\underset{\sim}{Y}_\alpha$, $\alpha = 1, 2, \ldots, n$ are each distributed normally with mean $\underset{\sim}{0}$ and covariance Σ, then a necessary and sufficient condition that</u>

$$B = \sum_{\alpha=1}^{n} \underset{\sim}{Y}_\alpha A \underset{\sim}{Y}_\alpha' = YAY'$$

<u>is distributed as $W(B \mid \Sigma, p, n)$ is that A is idempotent and of rank k where $p \leq k \leq n$, and where $Y = (\underset{\sim}{Y}_1, \underset{\sim}{Y}_2, \underset{\sim}{Y}_3, \ldots, \underset{\sim}{Y}_n)$.</u>

Proof: We shall first prove sufficiency. Since A is idempotent there exists an orthogonal matrix R, by Theorem A and I, such that

$$RAR' = \begin{bmatrix} I_k & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{0} \end{bmatrix}$$

Then

$$YAY' = YR'RAR'RY' = YR' \begin{bmatrix} I_k & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{0} \end{bmatrix} RY'.$$

Then define the (p x n) matrix Z such that $Z = YR'$, and partition Z and R such that $Z = (Z^*, Z^{**})$ where Z^* is a (p x k) matrix and let

$$R = [R^*, R^{**}] .$$

where R* is a (k x n) matrix. Then

$$YAY' = [Z^*, Z^{**}] \begin{bmatrix} I_k & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{0} \end{bmatrix} \begin{bmatrix} Z^{*'} \\ Z^{*'*} \end{bmatrix} = Z^* Z^{*'} .$$

The (1 x n) vector $\dot{\underset{\sim}{Y}}_i'$, (i = 1, 2, . . . , p) is distributed as $N(\underset{\sim}{0}, \sigma_{ii} I)$; because each element of $\dot{\underset{\sim}{Y}}_i'$ is an element from normal independent vectors with mean zero and variance σ_{ii}. Since z_{ij}, an element from Z, is a linear combination of the elements of $\dot{\underset{\sim}{Y}}_i'$ and the elements of $R^{*'}$ then z_{ij} is distributed as $N(0, \sigma_{ii} R_j' R_j)$. But $R_j' R_j = 1$ which implies that z_{ij} is distributed as $N(0, \sigma_{ii})$. Let the ijth element of R be ρ_{ij}. Then the

$$\begin{aligned} \operatorname{cov}(z_{ik}, z_{qr}) &= E\left\{[\dot{\underset{\sim}{Y}}_i' R_k - E(\dot{\underset{\sim}{Y}}_i' R_k)][\dot{\underset{\sim}{Y}}_q' R_r - E(\dot{\underset{\sim}{Y}}_q' R_r)]'\right\} \\ &= E\,[(\dot{\underset{\sim}{Y}}_i' R_k)(\dot{\underset{\sim}{Y}}_q' R_r)'] \\ &= E\,[\dot{\underset{\sim}{Y}}_p' R_k R_r' \dot{\underset{\sim}{Y}}_q] \\ &= E\,[\sum_t \sum_s y_{it}\, y_{qt}\, \rho_{tk}\, \rho_{sr}] \\ &= E\left[\sum_t y_{it}\, y_{qt}\, \rho_{tk}\, \rho_{tr} + \sum_{\substack{t \\ t \neq s}} \sum_s y_{it}\, y_{qs}\, \rho_{tk}\, \rho_{sr}\right] \end{aligned}$$

where

$$E(y_{rs}\, y_{tj}) = \begin{Bmatrix} \sigma_{rt} \text{ for all r and t when } s = j \\ 0 \text{ for all r and t when } s \neq j \end{Bmatrix} .$$

Therefore,

$$\operatorname{cov}(z_{ik}, z_{qr}) = \sum_t \sigma_{iq}\, \rho_{tk}\, \rho_{tr} + \sum_{\substack{t \\ t \neq s}} \sum_s (0)\, \rho_{tk}\, \rho_{sr}$$

$$= \sum_t \sigma_{iq} \rho_{tk} \rho_{tr} .$$

Since ρ_{tk} and ρ_{tr} are elements from rows of an orthogonal matrix, we then see that the cov $(z_{ik}, z_{qr}) = \sigma_{iq}$ if $k = r$ or the cov $(z_{ik}, z_{qr}) = 0$ if $k \neq r$. This then implies that Z_i^* is distributed as $N(\underset{\sim}{0}, \not\Sigma)$. Therefore, $B = Z^* Z^{*\prime} = YAY'$ is distributed as $W(B \mid \not\Sigma, p, n)$. Note that Z^* is a $(p \times k)$ matrix and therefore k, the rank of A, must be greater than or equal to p such that the original definition in Theorem J will be satisfied.

To prove necessity we shall compare the moment generating function of the Wishart distribution, which is, by Theorem M,

$$m_B(t) = \left| I - \not\Sigma t \right|^{-n/2}$$

with that of YAY'. We shall prove that A is idempotent if $B = YAY'$ is distributed as the Wishart distribution. We know there exists an orthogonal matrix R, by Theorem I, such that $RAR' = D = (d_{ii})$, a diagonal matrix. Then let $Z = YR'$ which implies that

$$YAY' = ZRAR'Z' = ZDZ' = \sum_{i=1}^{n} \underset{\sim}{Z}_i d_{ii} \underset{\sim}{Z}_i' = \sum_{i=1}^{n} d_{ii} \underset{\sim}{Z}_i \underset{\sim}{Z}_i' = Q$$

where $\underset{\sim}{Z}_i$ is a $(p \times 1)$ vector such that $\underset{\sim}{Z}_i$ is distributed as $NID(\underset{\sim}{0}, \not\Sigma)$ as indicated by the proof for sufficiency. Since rank of A is k, the moment generating function of

$$Q = \sum_{i=1}^{n} d_{ii} \underset{\sim}{Z}_i \underset{\sim}{Z}_i' \quad \text{is} \quad m_Q(t) = \left| I - D\not\Sigma t \right|^{-k/2}$$

Equating moment generating functions of B and Q we have

$$\left| I - D\Sigma t \right|^{-k/2} = \left| I - \Sigma t \right|^{-n/2}$$

which implies that k of the d_{ii}'s must be unity and the remaining d_{ii}'s must be zero. But the d_{ii}'s are, by Theorem I, the characteristic roots of A. Therefore, by Theorem A, the matrix A is idempotent.

Theorem 2. If the p variate independent vectors $\underset{\sim}{Y}_\alpha$, $\alpha = 1, 2, \ldots, n$ are distributed normally with mean $\underset{\sim}{\mu}_\alpha$ and covariance Σ^* then a sufficient condition that

$$B = \sum_{\alpha=1}^{n} \underset{\sim}{Y}_\alpha A \underset{\sim}{Y}_\alpha' = YAY'$$

be distributed as the non central Wishart distribution is that A be idempotent and of rank k where $p \leq k \leq n$. (Σ^* will be specified later).

Proof: Since A is idempotent there exists an orthogonal matrix, R, by Theorem I, such that

$$RAR' = \begin{bmatrix} I_k & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{0} \end{bmatrix}$$

Therefore, $B = YAY' = YR'RAR'RY' = Z^* Z^{*\prime}$, (where Z^* is specified in Theorem 1). Let $Z^* = YR^{*\prime}$, where $\overset{\circ}{\underset{\sim}{Y}}{}_i'$ is distributed $N(\underset{\sim}{\mu}_i j, \sigma_{ii} I)$, $i = 1, 2, \ldots, p$. Therefore, z_{ij}, the element in the i^{th} row and j^{th} column of Z^* ($z_{ij} = \overset{\circ}{\underset{\sim}{Y}}{}_i' R_j$) is distributed as $N(R_j' \underset{\sim}{\mu}_i j, \sigma_{ii} R_j' R_j)$. But z_{ij} is also distributed as $N(\underset{\sim}{\mu}_i R_j' j, \sigma_{ii})$ because R_j is a column from an orthogonal matrix which implies that $R_j' R_j = 1$. The

$$\text{cov}\,(z_{ik}, z_{qr}) = E\left\{[\dot{\underset{\sim}{Y}}_i' R_k - E(\dot{\underset{\sim}{Y}}_i' R_k)][\dot{\underset{\sim}{Y}}_q' R_r - E(\dot{\underset{\sim}{Y}}_q' R_r)]'\right\}$$

$$= E\left\{[\dot{\underset{\sim}{Y}}_i' R_k - \underset{\sim}{\mu}_i R_k' j][R_r' \dot{\underset{\sim}{Y}}_q - \underset{\sim}{\mu}_q j' R_r]\right\}$$

$$= E\ [\dot{\underset{\sim}{Y}}_i' R_k R_r' \dot{\underset{\sim}{Y}}_q' - \underset{\sim}{\mu}_q \dot{\underset{\sim}{Y}}_i' R_k j' R_r - \underset{\sim}{\mu}_i R_k' j R_r' \dot{\underset{\sim}{Y}}_q$$

$$+ \underset{\sim}{\mu}_i \underset{\sim}{\mu}_q R_k' j j' R_r] .$$

Now

$$R_k j' R_r = \sum_{s=1}^{n} \rho_{sr} \begin{bmatrix} \rho_{1k} \\ \rho_{2k} \\ \cdot \\ \cdot \\ \cdot \\ \rho_{pk} \end{bmatrix}$$

$$R_k' j R_r' = \sum_{s=1}^{n} \rho_{sk} (\rho_{1r}, \rho_{2r}, \ldots, \rho_{pr})$$

$$R_k' J R_r = \sum_{s=1}^{n} \rho_{sk} \sum_{j=1}^{n} \rho_{jr} .$$

Then

$$\text{cov}\,(z_{ik}, z_{qr}) = E\left\{\left[\sum_s y_{is}\, y_{qs}\, \rho_{sk}\, \rho_{sr} + \sum_{\substack{s \\ s \neq j}} \sum_j y_{is}\, y_{qj}\, \rho_{sk}\, \rho_{jr}\right] + \right.$$

$$\left[-\mu_q \sum_s y_{is}\, \rho_{sr}\, \rho_{sk} - \mu_q \sum_{\substack{s \\ s \neq j}} \sum_j y_{ij}\, \rho_{sr}\, \rho_{jk}\right] +$$

$$\left[-\mu_i \sum_s y_{qs} \rho_{sk} \rho_{sr} - \mu_i \sum_s \sum_{\substack{j \\ s \neq j}} y_{qj} \rho_{sk} \rho_{jr} \right] +$$

$$\left. \left[\mu_i \mu_q \sum_s \rho_{sk} \rho_{sr} + \mu_i \mu_q \sum_s \sum_{\substack{j \\ s \neq j}} \rho_{sk} \rho_{jr} \right] \right\}$$

$$= E \left\{ [1] + [2] + [3] + [4] \right\} .$$

Then for brackets 1, 2, 3, and 4, the expected value is, for [1], σ_{iq} if $k = r$ and zero if $k \neq r$, for [2], $-\mu_q \mu_i$ if $k = r$ and zero if $k \neq r$, for [3], $-\mu_i \mu_q$ if $k = r$ and zero for $k \neq r$, and for [4], $\mu_i \mu_q$ if $k = r$ and zero if $k \neq r$. Therefore, the cov $(z_{ik}, z_{qr}) = \sigma_{iq} - \mu_i \mu_q$ for $k = r$ and the cov $(z_{ik}, z_{qr}) = 0$ for $k \neq r$. These two statements imply that $\underset{\sim}{Z}_i^*$ is distributed $N(\underset{\sim}{\mu}_i^*, \Sigma^*)$ and $\underset{\sim}{Z}_i^*$ is distributed $NID(\underset{\sim}{\mu}_i^*, \Sigma^*)$. Therefore, $B = YAY' = Z^* Z^{*'}$ is distributed as the non-central Wishart where Σ^* and $\underset{\sim}{\mu}_i^*$ are specified as follows. The ijth element of Σ^* is $\sigma_{ij} - \mu_i \mu_j$, $i = 1, 2, \ldots, p$, $j = 1, 2, \ldots, p$. The elements in $\underset{\sim}{\mu}_i^*$ are $\mu_i R_j' j$, and the ijth element of T is

$$\sum_{s=1}^{k} \mu_s^2 R_i' J R_j$$

where $T = \underset{\sim}{\mu}^* \underset{\sim}{\mu}^{*'}$.

<u>Theorem 3. If $\underset{\sim}{Y}_i$ is distributed normally and independently, for all $i = 1, 2, \ldots, n$, with mean $\underset{\sim}{\mu}_i$ and variance Σ, then YAY' is</u>

<u>distributed as $W(B \mid p, n, \Sigma^*)$ if $[E(Y)]\, A\, [E(Y)]' = \underset{\sim}{0}$, and if A is idempotent of rank k where $p \leq k \leq n$.</u>

Proof: Since the matrix A is idempotent we know, by Theorems A and I, there exists an orthogonal matrix R such that

$$RAR' = \begin{bmatrix} I_k & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{0} \end{bmatrix} .$$

Then let $Z = YR'$ which implies that

$$\begin{aligned}
(EY)\, A\, (EY)' &= [E(ZR)]\, A\, [E(ZR)]' \\
&= (EZ)\, RAR'\, (EZ)' \\
&= (EZ) \begin{bmatrix} I_k & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{0} \end{bmatrix} (EZ)' \\
&= [E(Z^*,\ Z^{**})] \begin{bmatrix} I_k & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{0} \end{bmatrix} [E(Z^*,\ Z^{**})]' \\
&= (EZ^*)(EZ^*)'
\end{aligned}$$

From Theorem 2 we have shown that $\underset{\sim}{Z}_i^*$ is distributed $NID(\underset{\sim}{\mu}_i^*, \Sigma^*)$ which implies the expected value of each vector of Z^* is $E(\underset{\sim}{Z}_i^*) = \underset{\sim}{\mu}_i^*$ which is $\underset{\sim}{0}$ if $\underset{\sim}{\mu}_i^*$ is $\underset{\sim}{0}$. Therefore, YAY is distributed, by Theorem 1, as $W(B \mid p, n, \Sigma)$ if $\underset{\sim}{\mu}_i^* = \underset{\sim}{0}$.

<u>Theorem 4.</u> <u>If for all $\alpha = 1, 2, \ldots, n$, the vector $\underset{\sim}{Y}_\alpha$ has the p variate normal distribution with mean $\underset{\sim}{\mu}$ and variance covariance Σ, where $\underset{\sim}{Y}_i$ and $\underset{\sim}{Y}_j$ are independent for all $i \neq j$, then a sufficient condition that the forms YAY' and YBY' are independent is that</u>

<u>$AB = \underset{\sim}{0}$. (A and B are symmetric).</u>

Proof: Since $AB = \underset{\sim}{0}$, then $B'A' = \underset{\sim}{0}$ but $A = A'$ and $B = B'$, because of symmetry, therefore, $AB = BA = \underset{\sim}{0}$. Then, since A and B commute, there exists an orthogonal matrix R, by Theorem E, such that $RAR' = D_1$ and $RBR' = D_2$ where D_1 and D_2 are diagonal matrices. Also we have $RAR'RBR' = \underset{\sim}{0}$ because $R'R = I$ and $AB = \underset{\sim}{0}$, or $D_1D_2 = \underset{\sim}{0}$. If $D_1D_2 = \underset{\sim}{0}$, then the i^{th} diagonal element of D_1 being non zero implies, by Theorem E, that the i^{th} diagonal element of D_2 must be zero. Let

$$D_1 = \begin{bmatrix} D_{11} & \underset{\sim}{0} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{0} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{0} & \underset{\sim}{0} \end{bmatrix}$$

and

$$D_2 = \begin{bmatrix} \underset{\sim}{0} & \underset{\sim}{0} & \underset{\sim}{0} \\ \underset{\sim}{0} & D_{22} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{0} & \underset{\sim}{0} \end{bmatrix}$$

then make the transformation $Z = YR'$ which implies that

$$YAY' = ZRAR'Z' = ZD_1Z'$$

$$YBY' = ZRBR'Z' = ZD_2Z' \ .$$

Then YAY' depends only on the first s elements of Z, where D_{11} is

$s \times s$, also YBY' depends only on the $s + 1$ to r elements of Z where D_{22} is $(r-s) \times (r-s)$. In Theorem 2 we proved that $\underset{\sim}{Z}_i$ is distributed as $NID(\underset{\sim}{\mu}^*, \not\Sigma^*)$ which implies that YAY' and YBY' are independent if $AB = \underset{\sim}{0}$.

Theorem 5. Let $\underset{\sim}{Y}_\alpha$ be distributed $NID(\mu, \not\Sigma)$ $\alpha = 1, 2, \ldots, n$.

Assume the rank of $A_i > p$, and also assume that

$$\sum_{i=1}^{q} YA_iY' = YY'.$$

Then any two of the following conditions implies that YA_iY', $i = 1, 2, \ldots, q$, are distributed as independent non central Wishart distributions.

(1) $A_iA_j = \underset{\sim}{0}$ for all $i \neq j$.

(2) A_j is idempotent for all j.

(3) $\text{rank}(\Sigma A_i) = \Sigma(\text{rank } A_i)$.

Proof: Assume condition (1) and condition (2):

Condition (1) implies independence by Theorem 3 and condition (2) implies A_j in Wishart by Theorem 2.

Assume condition (1) and condition (3):

Condition (1) implies independence by Theorem 3 and condition (3) implies A_j is Wishart by Theorem H.

In fact condition (3) alone implies independent Wisharts. Assume condition (3) and condition (2):

Condition (2) implies Wisharts by Theorem 2 and condition (3) implies independence by Theorem H.

Again condition (3) alone implies independent Wisharts.

CHAPTER III

APPLICATIONS

Let us define the square of the vector

$$\underset{\sim}{Y}_i = \begin{bmatrix} x_i \\ y_i \end{bmatrix}$$

to be

$$\begin{bmatrix} x_i \\ y_i \end{bmatrix} (x_i \;\; y_i) = \begin{bmatrix} x_i^2 & x_i y_i \\ x_i y_i & y_i^2 \end{bmatrix} = \underset{\sim}{Y}_i \underset{\sim}{Y}_i' .$$

Then we have

$$\sum_{i=1}^{n} \underset{\sim}{Y}_i \underset{\sim}{Y}_i' = \begin{bmatrix} \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i y_i \\ \sum_{i=1}^{n} x_i y_i & \sum_{i=1}^{n} y_i^2 \end{bmatrix} .$$

Now the randomized complete block model

$$\underset{\sim}{Y}_{ij} = \underset{\sim}{\mu} + \underset{\sim}{\tau}_i + \underset{\sim}{\beta}_j + \underset{\sim}{e}_{ij}$$

where

$$i = 1, 2, \ldots, t$$

$$j = 1, 2, \ldots, b$$

where $\underset{\sim}{\tau}_i$ is a (p x 1) vector of treatments, $\underset{\sim}{\beta}_j$ is a (p x 1) vector of blocks, $\underset{\sim}{\mu}$ and $\underset{\sim}{e}_{ij}$ are (p x 1) vectors, can be written in matrix notation as

$$Y' = XB + e$$

where B is the matrix of parameters, μ, τ, and β and where Y' is an (n x p) matrix of randomally chosen responses, X is an (n x c) known matrix of fixed constants, B is a (c x p) matrix of unknown constants, and e is an (n x p) matrix such that $\underset{\sim}{e}_j$, an (n x 1) vector of random variables is distributed as NID($\underset{\sim}{0}$, Σ), for all j = 1, 2, . . . , p.

We know there exists a reparameterization on the model $Y' = XB + e$ to the model $Y' = Z\alpha + e$ such that $(Z'Z)\hat{\alpha} = Z'Y$ and such that $(Z'Z)^{-1}$ exists. The reparameterized model $Y' = Z\alpha + e$ may be written as $Y' = Z_1\alpha_1 + Z_2\alpha_2 + e$ where α_1 contains only the parameters $\underset{\sim}{\tau}_i$ while α_2 contains all other parameters. Then under the null hypothesis, that all treatment effects are the same, our model reduces to

$$Y' = Z_2\alpha_2 + e \, .$$

The Sums of Squares in the Analysis of Variance

Our analysis of variance table is then,

TABLE I

Analysis of Variance

Source	d.f.	S.S.	M.S.
Total	n	YIY'	
$R(\alpha)$	p	$\hat{\alpha}'Z'Y'$	
$R(\alpha_2)$	$p - r$	$\hat{\alpha}_2'Z_2'Y'$	
$R(\alpha_1 \mid \alpha_2)$	r	$\hat{\alpha}'Z'Y' - \hat{\alpha}_2'Z_2'Y' = B$	B/r
Error	$n - p$	$YIY' - \hat{\alpha}'Z'Y' = A$	$A/n - p$

where

$$\hat{\alpha} = (Z'Z)^{-1}Z'Y'$$

and

$$\hat{\alpha}_2 = (Z_2'Z_2)^{-1} Z_2' Y' .$$

Now

(3.1) $$B = YZ(Z'Z)^{-1} Z'Y' - YZ_2(Z_2' Z_2)^{-1} Z_2' Y'$$

$$= Y[Z(Z'Z)^{-1}Z' - Z_2(Z_2' Z_2)^{-1} Z_2']Y' .$$

Let

$$B_1 = I - ZS^{-1}Z'$$

and

$$B_2 = I - Z_2S_2^{-1}Z_2'$$

where

$$S^{-1} = (Z'Z)^{-1}$$

and

$$S_2^{-1} = (Z_2'Z_2)^{-1} .$$

Also we have

$$\begin{aligned} B_1B_1 &= (I - ZS^{-1}Z')(I - ZS^{-1}Z') \\ &= I - ZS^{-1}Z' - ZS^{-1}Z' + (ZS^{-1}Z')(ZS^{-1}Z') \\ &= I - 2ZS^{-1}Z' + ZS^{-1}Z' \\ &= I - ZS^{-1}Z' \end{aligned}$$

and

$$\begin{aligned} B_2B_2 &= (I - Z_2S_2^{-1}Z_2')(I - Z_2S_2^{-1}Z_2') \\ &= I - Z_2S_2^{-1}Z_2' - Z_2S_2^{-1}Z_2' + (Z_2S_2^{-1}Z_2')(Z_2S_2^{-1}Z_2') \end{aligned}$$

$$= I - 2Z_2S_2^{-1}Z_2' + Z_2S_2^{-1}Z_2'$$

$$= I - Z_2S_2^{-1}Z_2' \ .$$

Therefore B_1 and B_2 are idempotent matrices. Then

$$(B_2 - B_1)(B_2 - B_1) = B_2^2 - B_2B_1 - B_1B_2 + B_1^2$$

$$= B_2 - B_2B_1 - B_1B_2 + B_1 \ .$$

Now

$$B_2B_1 = (I - Z_2S_2^{-1}Z_2')(I - ZS^{-1}Z')$$

$$= I - ZS^{-1}Z' - Z_2S_2^{-1}Z_2' (I - ZS^{-1}Z')$$

but

$$Z'(I - ZS^{-1}Z') = Z' - Z'ZS^{-1}Z'$$

$$= Z' - Z' = \underset{\sim}{0}$$

therefore

$$\begin{bmatrix} Z_1' \\ Z_2' \end{bmatrix} (I - ZS^{-1}Z') = \underset{\sim}{0}$$

so

$$B_2B_1 = B_1$$

also

$$B_1B_2 = (I - ZS^{-1}Z')(I - Z_2S_2^{-1}Z_2')$$

$$= I - ZS^{-1}Z' - Z_2S_2^{-1}Z_2'(I - ZS^{-1}Z')$$

$$= I - ZS^{-1}Z'.$$

Therefore,

$$B_1B_2 = B_1$$

so

$$(B_2 - B_1)(B_2 - B_1) = B_2 - B_1 - B_1 + B_1$$

$$= B_2 - B_1$$

which implies that $B_2 - B_1 = B$, where B is defined in equation (3.1), is idempotent which then implies that YBY' is distributed $W(B \mid \not{\Sigma}, p, n)$ by Theorem 1, if $\underset{\sim}{Y}_i$ is distributed NID($\underset{\sim}{0}$, I).

We shall now show that A is Wishart.

$$A = YIY' - \hat{\alpha}'Z'Y' = YIY' - YZS^{-1}Z'Y' = Y(I - ZS^{-1}Z')Y'$$

but $I - ZS^{-1}Z'$ is idempotent in the above paragraph. Therefore, A is distributed $W(A \mid \not{\Sigma}, p, n)$ by Theorem 1, if $\underset{\sim}{Y}_i$ is distributed NID($\underset{\sim}{0}$, $\not{\Sigma}$).

Now A and B are independent Wisharts, by Theorem 3, if

$B_1(B_2 - B_1) = \underset{\sim}{0}$,

$$\begin{aligned} B_1(B_2 - B_1) &= (I - ZS^{-1}Z')(ZS^{-1}Z' - Z_2S_2^{-1}Z_2') \\ &= ZS^{-1}Z' - Z_2S_2^{-1}Z_2' - ZS^{-1}Z'ZS^{-1}Z' + ZS^{-1}Z'Z_2S_2^{-1}Z_2' \\ &= ZS^{-1}Z' - Z_2S_2^{-1}Z_2' - ZS^{-1}Z' + ZS^{-1}Z'Z_2S_2^{-1}Z_2' \\ &= (ZS^{-1}Z' - I)Z_2S_2^{-1}Z_2' \end{aligned}$$

but

$$\begin{aligned} (ZS^{-1}Z' - I)Z &= ZS^{-1}Z'Z - Z \\ &= Z - Z \\ &= \underset{\sim}{0}. \end{aligned}$$

Therefore,

$$(ZS^{-1}Z' - I)(Z_1, Z_2) = \underset{\sim}{0}$$

so

$$(ZS^{-1}Z' - I)Z_2 = \underset{\sim}{0}$$

which implies by Theorem 3 that A and B are independent.

Test Function

From (1) we know that

$$\frac{|A|}{|A + B|}$$

is distributed as $U_{2, q, r}$ if $\underset{\sim}{Y}_i$ is a (2×1) vector, where A is the error sum of squares with r degrees of freedom, and where B is treatment sum of squares with q degrees of freedom. We also know from (1) that

$$v = \frac{1 - \sqrt{U_{2, q, r}}}{\sqrt{U_{2, q, r}}} \left(\frac{r - 1}{q} \right)$$

is distributed as $F_{2q, 2(r-1)}$ which we shall use as our test function.

Example

We shall present a set of hypothetical data and test the null hypothesis that there is no difference between treatments, i.e.

$$\begin{bmatrix} \tau_{11} \\ \tau_{12} \end{bmatrix} = \begin{bmatrix} \tau_{12} \\ \tau_{22} \end{bmatrix} = \ldots = \begin{bmatrix} \tau_{1n} \\ \tau_{2n} \end{bmatrix}.$$

In this example we shall have four treatments, ten blocks with two measurements on each treatment in every block. For example, we might like to measure height and weight of twenty one year old males in the United States. We shall assume the randomized complete block model

$$\underset{\sim}{y}_{ij} = \underset{\sim}{\mu} + \underset{\sim}{\tau}_i + \underset{\sim}{\beta}_j + \underset{\sim}{e}_{ij}.$$

The statistical layout of this data is given in Table II and the covariance analysis is given in Table III.

TABLE II

Data

Blocks \ Trts.	1	2	3	4	Totals
1	32.83 88.63	16.70 79.57	14.72 59.61	9.16 40.61	73.41 268.42
2	23.18 69.66	31.19 66.88	14.64 60.29	12.06 42.98	81.07 239.81
3	40.82 54.72	30.84 59.71	9.68 46.54	8.16 40.96	89.50 201.93
4	27.22 51.45	17.88 72.93	13.48 39.12	6.96 49.07	65.54 212.57
5	21.89 45.13	34.21 68.77	12.79 47.41	10.45 53.75	79.34 215.06
6	25.74 67.19	39.44 71.31	44.82 54.93	7.92 48.74	117.92 242.17
7	22.64 78.79	52.88 62.81	16.03 57.30	9.18 42.65	100.73 241.55
8	21.20 75.76	34.69 67.74	26.61 72.20	10.77 48.44	93.27 264.14
9	30.05 52.73	32.71 63.24	19.15 46.43	12.93 42.38	94.84 204.78
10	27.31 65.04	31.32 65.89	21.37 62.15	8.87 45.15	88.87 238.23
Totals	272.88 649.10	321.86 678.85	193.29 545.98	96.46 454.73	884.49 2328.66

TABLE III

Covariance Analysis

Source	d.f.	Σx^2	Σxy	Σy^2	d.f.	$\Sigma d^2_{y.x}$
Total	39	5145.47	2902.42	6180.29		
Blocks	9	493.29	90.77	1238.99		
Treatments	3	2914.23	3009.77	3137.73		
Error	27	1737.95	-198.12	1803.56	26	1780.98
Error plus Trts.	30	4652.18	2811.65	4941.29	29	3242.01

Let us note at this point that

$$|A| = \begin{vmatrix} \Sigma x_e^2 & \Sigma xy_e \\ \Sigma xy_e & \Sigma y_e^2 \end{vmatrix}$$

$$= \Sigma x_e^2 \, \Sigma y_e^2 - (\Sigma xy_e)^2$$

$$= \Sigma x_e^2 \left[\Sigma y_e^2 - \frac{(\Sigma xy_e)^2}{\Sigma x_e^2} \right]$$

$$= \Sigma x_e^2 \, \Sigma d^2_{y.xe}$$

where A is the error sum of squares. Likewise we have

$$\left|A + B\right| = \Sigma x_{te}^{2} \, \Sigma d_{y.xte}^{2}$$

where B is the treatment sum of squares and where the subscripts e and te denote error and treatment plus error sum of squares respectively. Therefore, we have

$$\left|A\right| = (1780.98)(1737.95)$$

$$= 3,095,256.72$$

and

$$\left|A + B\right| = (3242.01)(4652.18)$$

$$= 15,082,415.54 \; .$$

Then

$$\frac{\left|A\right|}{\left|A + B\right|} = 0.2052$$

which gives us

$$v = \frac{1 - \sqrt{0.2052}}{\sqrt{0.2052}} \left(\frac{26}{3}\right) = 10.4650 \; .$$

The tabulated F value at the 95% level with 6 and 52 degrees of freedom is 2.39. Comparing our v value with this F value we then shall reject the null hypothesis, there is no difference in treatments, at the 95% level.

CHAPTER IV

SUMMARY AND CONCLUSIONS

Suppose we have two characteristics that are to be tested under the hypothesis that all treatment effects are equal for the first characteristic and that all treatment effects are equal for the second characteristic. If we choose to test these hypotheses using the univariate model then test $H_o^{(1)}: \tau_1^{(1)} = \tau_2^{(1)} = \ldots = \tau_q^{(1)}$ and test $H_o^{(2)}: \tau_1^{(2)} = \tau_2^{(2)} = \ldots = \tau_q^{(2)}$ each as a single test. If we choose to test these hypotheses using the multivariate model then test $H_o^{(1)}$ and $H_o^{(2)}$ simultaneously, that is, our single test and hypothesis shall be H_o:

$$\begin{bmatrix} \tau_1^{(1)} \\ \tau_1^{(2)} \end{bmatrix} = \begin{bmatrix} \tau_2^{(1)} \\ \tau_2^{(2)} \end{bmatrix} = \ldots = \begin{bmatrix} \tau_q^{(1)} \\ \tau_q^{(2)} \end{bmatrix} .$$

In the univariate model we shall reject, on the average, $H_o^{(1)}$ 100α % of the time when $H_o^{(1)}$ is true and we shall reject, on the average, $H_o^{(2)}$ 100α % of the time when $H_o^{(2)}$ is true. When $H_o^{(1)}$ is rejected, for example, we shall imply there exists at least one inequality in $H_o^{(1)}$. In the multivariate model we shall reject, on the average, H_o 100α %

of the time when H_o is true and when H_o is rejected, this will imply there exists at least one inequality in $H_o^{(1)}$ or in $H_o^{(2)}$ or there exists at least one inequality in both $H_o^{(1)}$ and $H_o^{(2)}$.

In the univariate model let us assume $H_o^{(1)}$ and $H_o^{(2)}$ are both true and let us denote the probability of rejection of $H_o^{(1)}$ by A and the probability of rejection of $H_o^{(2)}$ by B. Then $Pr(A \text{ or } B) = Pr(A) + Pr(B) - Pr(AB)$. Let us first assume that $H_o^{(1)}$ and $H_o^{(2)}$ are independent. We then have $Pr(A \text{ or } B) = Pr(A) + Pr(B) - Pr(A)Pr(B) = \alpha\% + \alpha\% - (\alpha\%)^2 = 1 - (1 - \alpha\%)^2$. Next, assuming that $H_o^{(1)}$ and $H_o^{(2)}$ are functionally related we have $Pr(A \text{ or } B) = Pr(A) + Pr(B) - Pr(A|B) = \alpha\% + \alpha\% - \alpha\% = \alpha\%$. We see then that, Pr(rejection in multivariate model) $= \alpha\% \leq$ Pr(rejection in univariate model) $\leq 1 - (1 - \alpha\%)^2$. Note that the probability of rejection in the multivariate model is at least as small as the probability of rejection in the univariate model. Also if m hypothesis are being tested instead of two, Pr(rejection in multivariate model) $=$ $\alpha\% \leq$ Pr(rejection in univariate model) $\leq 1 - (1 - \alpha\%)^m$, which implies, when our hypotheses are true and when m is large that the probability of rejection in univariate model approaches one.

If N experiments are run with m characteristics in each experiment, where N is large, then our error rate per decision is the same in the univariate model as in the multivariate model, remembering that in the univariate model we make m decisions in each of the N experiments while in the multivariate model we make only one decision for each experiment. Our error rate is,

$$\frac{\text{expected number of wrong decisions}}{\text{number of decisions}} = \frac{\alpha\%N + \ldots + \alpha\%N}{mN} = \alpha\%$$

Our error rate per experiment, in the univariate model is,

$$\frac{\text{expected number of wrong decisions}}{\text{number of experiments}} = \frac{\alpha\%N + \ldots + \alpha\%N}{N} = m\alpha\% .$$

We then see that if we want the error rate per experiment to be $\alpha\%$ in the univariate model we must test each decision at the $\alpha\%/m$ level.

Although the power of the test is not considered in this paper, it seems that if we want to control error rate per experiment, we should test as in the multivariate model if m is appreciable. If we want to control error rate per decision it seems that the univariate test might be best.

Suggestions For Future Study

It would be interesting to evaluate the power of these two tests such that they could be compared at greater length.

BIBLIOGRAPHY

(1) Anderson, T. W. Introduction to Multivariate Statistical Analysis, New York: John Wiley and Sons, Inc., 1958.

(2) Graybill, Franklin A. and Marsaglia, George. "Idempotent Matrices and Quadratic Forms in the General Linear Hypothesis." Annals of Mathematical Statistics, XXVIII (September, 1957), 678-86.

(3) Graybill, Franklin A. General Linear Hypothesis Theory. Unpublished Notes, 1959.

(4) Hartley, H. O. "Some Recent Developments in Analysis of Variance." Communications on Pure and Applied Mathematics, VIII (February, 1955), 47-72.

VITA

Del Lon West
candidate for the degree of
Master of Science

Thesis: ON THE MULTIVARIATE ANALYSIS OF VARIANCE

Major: Statistics

Minor: Mathematics

Biographical and Other Items:

Born: October 21, 1934 at Ryan, Oklahoma

Undergraduate Study: Oklahoma University, 1952-54
Southeastern State College, 1954-57

Graduate Study: Oklahoma State University, 1957-59

Experiences: Student and Graduate Assistant in the Mathematics Department, 1957-59

Member of The Institute of Mathematical Statistics

Date of Degree: May 24, 1959